个人信息大数据与刑事正当程序的冲突及其调和

裴　炜*

内容提要：信息革命引发现代国家治理发生相应变革，这集中体现在国家权力与公民权利的互动上。在此背景下，个人信息保护不仅针对信息本身，还应防止因个人信息被滥用而侵害公民的合法权益。具体到刑事司法领域，以个人信息为基础的大数据在介入犯罪治理活动后，其具有的过程性和算法依赖性、以行为模式为前提假设的预测性、基于数据挖掘的新认知范式和数据碎片性等特性，引发犯罪治理思路和模式的相应转变，这尤其表现在两个方面：一是服务于刑事诉讼的数据收集、存留及共享义务的扩张，二是风险防控思维下犯罪治理活动启动时点的前移。个人信息大数据在助力犯罪风险评估从而优化刑事司法资源配置的同时，亦与刑事正当程序发生冲突，其中又以无罪推定原则、控辩平等原则和权力专属原则为甚。鉴于社会信息化的总体趋势，要调和这些冲突，需要以信息革命引发的“权力—权利”二元互动关系变革为出发点，寻求犯罪控制与保障人权两项刑事司法基本价值之间的新平衡点，并对具体的程序规则进行修正。

关键词：个人信息　大数据　犯罪治理　正当程序　比例原则

自上世纪50年代起，社会活动通过网络媒介不断数字化、信息化。而在大数据技术得到广泛应用的背景下，个人信息的搜集、存储、分析、使用，在成为社会治理重要手段的同时，也面临着被不当使用的风险。具体到犯罪治理领域，一方面，个人信息大数据在打击新型网络犯罪方面开始发挥越来越重要的作用；另一方面，这种新型犯罪治理手段也在更深的层次上撼动着传统刑事司法的运行规律和基本模式。更重要的是，这种变动不仅体现在以事实真相为依托的犯罪控制层面，也体现在以保障公民基本权利为核心的正当程序层面。

围绕个人信息大数据在刑事司法领域的具体应用，目前主要存在两种声音：一种是不

* 北京航空航天大学法学院副教授。

本文为北京市社会科学基金青年项目“网络犯罪电子证据原理探析与规则构建”（16FXC026）、司法部中青年项目“比例原则视域下个人信息保护的刑事司法规则研究”（17SFB3024）成果。

断强调其在强化犯罪控制方面所蕴含的巨大潜力，〔1〕另一种则是持续警示其在运用过程中与基本权利保护的紧张关系。就后者而言，研究重点主要在于对包括隐私权在内的个人信息权等实体性权利的保障，而对于程序性权利学界尚未有足够关注。

本文从刑事司法探求真相以打击犯罪、遵循正当程序以保障人权的基本任务出发，在国家刑罚权与公民基本权利互动的理论框架下，剖析个人信息大数据对刑事正当程序构成的挑战，以及立法者应如何修订具体规则以应对这些挑战。首先，第一部分围绕个人信息的使用与保护，就信息革命背景下“权力—权利”二元互动的新特征进行分析。其次，第二部分就大数据的含义、特质及其在犯罪治理中的具体应用进行论述。再次，以大数据的现实应用为基础，进一步分析这些应用与正当程序之核心要求的冲突，其中主要关注无罪推定原则、控辩平等原则和权力专属原则。最后，探讨立法者应如何在平衡犯罪控制与保障公民基本权利的过程中，应对这些冲突。

一、基于个人信息的“权力—权利”二元互动

在刑事司法领域，法律一方面确认公安司法机关侦查、起诉、审判、惩罚犯罪的权力，另一方面通过确认犯罪嫌疑人、被告人的程序性权利来划定刑罚权的合理边界。而由于信息革命，社会的数字化、网络化使得传统法律构建的“权力—权利”制衡关系面临新的挑战。一方面，以包括隐私在内的公民个人信息为中心而展开的权力与权利的角斗不断升温；另一方面，在这场角斗中，原本身处二元格局之外、包括网络中介在内的第三方主体的介入不断强化。

在这一新格局中，各方的互动呈现出多种形式，且彼此之间存在张力。例如：用户依靠集合化的评论对网络服务提供商施加压力，同时其个人信息亦由网络服务提供商搜集，用户受其引导，甚至用户由此养成新的消费习惯；国家通过网络和大数据强化犯罪控制能力，而这种能力很难与泛化的社会监控相剥离，与此同时，犯罪能力的强化也发生在犯罪领域；网络中介出于经营目的或法律规定，承担着维护用户个人信息的职责，但同时也负有配合国家执行特定行政或司法任务的义务。

毫无疑问，通过信息这一媒介而发生的国家权力与公民权利互动模式的变革，将随着社会数字化、网络化的不断深入而继续下去，进而对法律规定的修正、补充乃至重构产生影响。在进一步探讨个人信息大数据对刑事诉讼基本原则和制度构成的挑战之前，需要处理三个层层递进的前提性问题：第一，如何理解信息革命背景下的“信息”；第二，国家权力在社会信息化的背景下会发生何种变化；第三，公民权利在该背景下会发生何种变化。

（一）信息革命与信息的质变

总结各类文献中的相关表述，“信息”的定义大致可分为两种：描述性定义和功能性定义。前者侧重于信息的属性，例如强调信息反映事物的形成、关系和差别，即“信息就是差异”。〔2〕后者则主要着眼于构成“信息”之事物的功能性价值，例如将“消除随机不确定性”作为“信息”定义的核心。〔3〕但无论采用何种定义，一般认为信息至少包含以下两

〔1〕 See Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, Mich. St. L. Rev., 2016, pp. 947 – 1017.

〔2〕 See Giuseppe Longo (ed.), *Information Theory: New Trends and Open Problems*, Springer Vienna, 1975, p. 3.

〔3〕 See C. E. Shannon, *A Mathematical Theory of Communication*, 27 The Bell System Technical Journal 379 – 423 (1948).

个属性。首先，信息不同于“数据”和“知识”。信息是经过加工且有意义的数据，知识则是发展成熟且真实的信息。〔4〕“数据挖掘”（data mining）正是基于信息与数据的区分而形成的概念，这一区分也指示出挖掘的目的即获取信息。其次，信息具有可为人类理性认知和理解的意义。信息意义的形成，一方面依赖理解该意义的主体所处的特定时期、环境和背景，另一方面则取决于该主体的目的、逻辑和经验。在以上内外两重因素的共同作用下，作为信息形成基础的数据被筛选、分类和结构化。

基于上述两个属性，“信息”的定义可以大致抽象为“数据 + 意义”这一公式。符合该公式的信息自古就存在，但信息真正成为人类社会发展的关键力量，或者说引发所谓信息革命，则始于“二战”之后。信息革命的核心在于，信息开始被赋予传统的工具意义之外的角色，信息可以演化为资源、商品、财产、中介甚至是社会建构力（constitutive force）。〔5〕与之相对应，信息自身的价值与其内容价值发生剥离，其本身成为组织管理或社会治理的对象。例如信息生命周期管理理论，即着眼于信息“从生到死”的各个流程进行系统性分析，以及在该系统下对特定类型的信息进行管理规则的建构。〔6〕

在信息革命初期，就有学者预言信息管理及信息价值的实现将面临三重挑战：其一是基于信息量体量的技术性挑战；其二是针对含义和真实性的语义学挑战；其三是基于该含义与真实性所形成的针对人类行为模式的影响性挑战。〔7〕随着全球数字化、网络化的不断深入，上述三重挑战已不仅限于事物的局部或孤立的个体或事件，而是演变成社会整体治理中不可回避的事实。特别是在全面进入大数据时代之后，社会面临的问题已经不单是技术层面的问题，而更在于伦理道德层面，即如何理解这种技术对于公民、社会组织以及政府在权利义务规则方面产生的深远影响。个人信息在成为公民基本权利对象的同时，也是国家权力行使的重要资源和媒介，两者之间的张力在“权力—权利”二元互动最为典型和直接的刑事司法领域被进一步强化。〔8〕

（二）信息革命背景下的权力演化

在社会治理的语境下，权力概念总是涉及权力行使者与权力行使对象，以描述两者之间的一种特殊关系。〔9〕通过行使权力，权力对象的行为可能向着更符合权力主体所欲之方向或效果发展。从上述“权力主体—权力对象”的关系出发，社会心理学家提出了权力的五种基础：强制、奖励、正当性、专业性、集体参照。〔10〕有学者进一步将信息视为社会权

〔4〕 关于数据、信息、知识三者之间的关系，参见 Luciano Floridi, *Information: A Very Short Introduction*, Oxford University Press, 2010, pp. 21 – 25。

〔5〕 See Sandra Braman, *Change of State: Information, Policy, and Power*, The MIT Press, 2006, pp. 11 – 12.

〔6〕 2016 年，全球网络存储工业协会（Storage Networking Industry Association, SNIA）更新了信息生命周期管理的定义，即在信息自产生直至消灭的完整周期中，围绕信息价值的优化和成本效率分析而形成的一系列政策、过程、操作、服务和工具。See SNIA, *The 2016 SNIA Dictionary*, available at https://www.snia.org/education/dictionary, visited on August 4, 2017.

〔7〕 See Warren Weaver, *The Mathematics of Communication*, 181 (1) Scientific American 11 (1949).

〔8〕 戴维·加兰德认为，信息技术的发展使得原本不属于政策管理的事项进入公共政策领域，进而对社会资源在犯罪领域的分配产生影响。See David Garland, *The Culture of Control: Crime and Social Order in Contemporary Society*, The University of Chicago Press, 2001, pp. 18 – 19.

〔9〕 See Robert A. Dahl, *The Concept of Power*, 2 (3) Systems Research and Behavioral Science 202 – 204 (1957).

〔10〕 See French and Raven, *The Bases of Social Power*, in Dorwin Cartwright (ed.), *Studies in Social Power*, University of Michigan Press, 1959, pp. 150 – 167.

力的第六种基础，认为一方主体通过占有信息并控制另一方主体获取信息的渠道和程度，同样可以构成前者的权力来源。〔11〕

信息之所以能够成为独立的权力基础，是因为如何处理信息是个人认知、判断和行为的前提。一方面，信息作为一种社会资源，可以促成“权力关系的非对称性”，〔12〕即可以“不顾参与该行为的其他人的反抗而实现自己的意志”。〔13〕另一方面，在社会数字化、网络化的背景下，信息对权力的基本属性和实现形式形成了区别于其他权力基础的独特影响。就前一方面而言，权力的运行存在着一种从潜在状态发展为现实状态的过程；这种潜在状态不仅意味着使用已知的权力资源或技术的可能性，还意味着基于权力主体占有的其他资源生成新的权力资源或技术的可能性。在信息性权力的语境下，信息并不必然直接构成权力本身，而是更多扮演使权力得以实现和有效运行的能源和催化剂，这是信息作为软性权力的核心特征。就后一方面，即信息之于权力实现形式的具体作用而言，这种催化效果可以从权力关系的广延性、综合性和强度这三个维度加以分析。“广延性”用于描述权力对象的规模，即权力的行使是只涉及特定社会组织中的部分对象还是全部对象；“综合性”用于描述权力所涉及的事务的类型和范围；“强度”则用于描述权力影响权力对象的行为的程度。

首先，社会信息化主要从权力的物理边界和权力的执行主体这两个层面扩展了权力的广延性。现实世界的物理边界是权力广延性的重要限制因素之一，而网络革命或者数字革命的一大特征就是，物理边界之于社会治理的意义开始弱化；国家权力得以充分触及主权辖区内的各个角落，从而前所未有地扩展了权力的广延性。就权力的执行主体而言，数字技术为社会规范的执行提供了相对便宜的手段，原先国家权力机关对规范执行的垄断藉由同意或者协商而被打破和稀释。〔14〕在这个意义上，执行主体的分散化在一定程度上使权力延伸至更为广阔和细微的社会层面。

其次，从权力的综合性的视角看，同样可以观察到信息化社会带来的影响，这种影响更多依赖认知范式的转变。人类历史上处理数据的范式有经验范式、理论范式和计算范式，而数字时代的新范式是数据挖掘（data exploration）范式。〔15〕相对于前三种范式，数据挖掘范式最典型的特征是，其转变了传统的围绕特定认知对象或假设而进行的数据搜集模式，取而代之以基于广泛、全面、深度的数据搜集而形成认知对象或假设的过程。通过此种范式转变，权力行使伊始所针对的领域及对象变得模糊，跨界的数据搜集和共享成为常态，权力运行可能辐射的事项范围也因难以事前预测而呈现出泛化的趋向。

最后，信息社会对权力的影响也反映在权力的强度层面。对此，可从信息输入及输出这两个角度进行分析。就信息输入而言，不同主体之间因信息输入的质与量上的差异而形

〔11〕 See B. H. Raven, *Power and Social Influence*, in Ivan Dale Steiner & Martin Fishbein (eds.), *Current Studies in Social Psychology*, Holt, Rinehart and Winston, 1965, pp. 127 – 145.

〔12〕 See Dannies H. Wrong, *Power: Its Forms, Bases, and Uses*, Transaction Publishers, 1995, pp. x – xiii.

〔13〕 See Max Weber, *Economy and Society*, Guenther Roth and Claus Wittich (eds.), Bedminster Press, 1968, Volume one, p. 53.

〔14〕 See Lawrence Lessig, *Code Version 2.0*, Basic Books, 2006, pp. 145 – 149.

〔15〕 See Randal E. Bryant, Randy H. Katz & Edward D. Lazowska, *Big-data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society*, issued December 22, 2008, available at http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/Big_Data.pdf, visited on December 20, 2016.

成在互动过程中的不平等地位。就信息输出而言，信息作用于权力的强度，表现为权力主体通过对数据进行选择性收集、筛选、拼组以形成信息，从而在此基础上强化其说服权力行使对象的能力。

从世界范围来看，信息在权力运行中的角色随着上世纪福利国家和风险社会理论的发展而不断强化，国家权力的运行和实现形式呈现出以下四个主要特征：首先，国家越来越少成为执法行为的唯一提供者，执法活动越来越多采用“分包”的形式进行，〔16〕私主体逐渐转变为权力中介。其次，在权力中介的生成与参与下，国家权力呈现出外溢的特征，传统意义上的公私边界逐渐模糊。〔17〕再次，在公私边界模糊的背景下，围绕着作为信息基础的数据收集、分析和共享，政府机构之间以及机构与中介之间的网状结构不断强化。最后，这一信息网络的构建逐步融入社会机理，以开放式、扩散式的生长方式取代原先针对特定目标或对象而实施的信息收集活动。

（三）基于个人信息的权利应对

当信息被定义为权力基础而作用于社会时，其产生的效果不仅是规则体系发生变化，更在于原先以宪法为核心而构建的基本权利体系以及相应原则出现适用上的困难。〔18〕在信息革命兴起之时，这种困难首先表现为隐私权与言论自由之间的冲突。彼时基于信息而产生的规则冲突，其核心关注点是权利位阶划分及异类权利之间的衡平，国家权力更多扮演调停者的角色。上世纪七八十年代之后，权利体系的对应方逐渐转变为系统性、大规模的数据搜集和存留行为，隐私权概念难以应对这类新的现象，政府公共行为也开始成为权利互动的另一方关键主体。21 世纪之后，恐怖主义活动和严重、复杂犯罪活动的升级、变异，使得“权力—权利”二元互动进一步复杂化：一方面，围绕个人信息保护形成了包括被遗忘权在内的一系列新型权利；另一方面，为实现高效的犯罪治理，强化国家权力的呼声也在高涨。

随着权力呈现出新的特质，与之伴生的权利概念也开始发生变化。从个人信息保护的角度来看，这种变化在权利的主体、性质、客体、对应或义务主体（respondent）、正当性基础或依据等五个方面〔19〕均有所体现。

第一，就权利主体而言，数据以及在数据基础上形成的信息，其权属并不清晰。〔20〕这种模糊性既体现在个人信息作为分析的原始数据之时，也体现在其作为分析的结果之时；既涉及存储于特定载体的静态数据，也涉及实时传输和处理中的动态数据。特别是在大数据的背景下，数据的权属是基于存储、处理而产生，还是基于相关性而产生，以及在此基础上，“个人信息”之“个人”描述的是信息客体还是主体，围绕这些问题目前均未形成较为合理的权利体系；由此进一步引发实践中针对个人信息收集、存储、分析、推断、使用、交易等行为的合法性争议。〔21〕

---

〔16〕 See Eric Brousseau, *et al.* (eds.), *Governance, Regulations and Powers on the Internet*, Cambridge University Press, 2012, pp. 5 – 6.

〔17〕 同上书，第 1 页以下。

〔18〕 参见前引〔5〕，Braman 书，第 39 页。

〔19〕 See Alan Gewirth, *Human Rights: Essays on Justification and Applications*, University of Chicago Press, 1982, p. 2.

〔20〕 参见龙卫球 《数据财产权构建及其体系研究》，《政法论坛》2017 年第 4 期，第 64 页以下。

〔21〕 See Ali M. Al-Khouri, *Data Ownership: Who Owns "My Data"?*, 2 (1) International Journal of Management & Information Technology 1 – 8 (2012).

第二，就权利的性质而言，个人信息属于财产权、人格权亦或混合型新型权利的争论，随着网络革命的深入而愈演愈烈；〔22〕其背后则是个人信息难以嵌入以民事权利体系为基本模型、以侵权损害赔偿为主要救济手段、以实害而非风险为导向的传统法律框架这一现实。

第三，隐私权得以成为一项基本权利，但其所依赖的价值基础开始被反思。传统上认为隐私权根源于人与人之间的隔离，而这种隔离造成了人与人之间的差异，这使得人们得以进行创造性活动并得出不同的观点，进而形成不同的生活方式，尊重多元化和自由的社会也由此产生。〔23〕这种价值基础变异的一个显著表现是，在技术语境下描述的数据保护与法律或道德层面的隐私出现了语义上的偏差。〔24〕

第四，在反思隐私权的基础上，权利客体呈现出多元化的发展趋势，衍生出包括被遗忘权、可携带权等在内的新的权利概念。相对于传统的隐私权，新型权利的一个共同特征是，其关注点由实害转向风险。这种转变直接带来个人信息保护中的“隐私困境”，即一方面，潜在的侵害风险使得公民的个人信息保护诉求不断增多和复杂化，直观地表现为侵犯个人信息的行为界定由非法交易、披露等行为向搜集、存留等行为扩张，权利边界呈现出模糊化的趋势。另一方面，基于风险形成的侵害结果的间接性和延迟，使得个人识别和保护其信息的意识难以与其权利诉求相匹配。换言之，人们在呼吁保护个人信息时，自己往往首先成了个人信息的泄露者。

第五，如果说在传统的隐私权语境下，权利主体在隐私诉求与自我保护能力之间还可以保持一定的同一性，那么，当前基于切实的技术和分析能力上的差异，呈现出权利诉求者与权利保护者相剥离的发展趋势，〔25〕后者转归于国家、网络平台等主体。

综上，在信息质变为权力基础的当下，国家权力与公民权利的二元互动呈现出紧张而又相互依赖的关系。围绕着个人信息保护与使用而展开的争论，是这种关系最为典型的表现。一方面，信息革命带来的风险化社会使得权力和权利双方的诉求均呈现扩张趋势，这集中体现在界定个人信息时规则的游移不定，无论这种界定是着眼于“可识别性”还是“同意”。这种双向扩张不可避免地造成特定领域的紧张态势，这特别表现在个人信息权与国家或社会公共安全之间的竞争关系。另一方面，权利诉求的扩张并未同步强化权利主体实现该诉求的能力，相反，信息革命使得这种能力无论在识别侵害风险方面，还是在有效救济方面，均不断弱化，从而强化了权利主体对于国家基于公权力进行保护和救济的依赖。具体到刑事司法领域，权力与权利之间紧张却相互依赖的关系，在个人信息大数据的背景下尤为明显。下面本文以大数据介入犯罪治理活动为切入点，对这种关系作进一步的审视。

二、犯罪治理活动中个人信息大数据的介入

个人信息大数据介入犯罪治理活动，一方面与社会生活整体的数字化、网络化密切相

〔22〕 这种争论尤其体现在民法总则制定过程中对“个人信息”的权利安置及表述上。参见《民法总则立法背景与观点全集》编写组编《民法总则立法背景与观点全集》，法律出版社2017年版，第9页，第18页，第24页，第53页，第89页。

〔23〕 See Timothy Macklem, *Independence of Mind*, Oxford University Press, 2008, p. 56.

〔24〕 See Philip Agre and Marc Rotenberg (eds.), *Technology and Privacy: The New Landscape*, MIT Press, 1997, pp. 39 – 40.

〔25〕 参见前引〔8〕，Garland书，第78页。

关；另一方面，就具体的介入方式和产生的影响而言，则受制于大数据的基本特性。以下从大数据的基本特性入手，基于其引发的认知范式转变，进一步分析大数据在犯罪治理活动中的主要介入路径。

（一）认知范式转变背景下的大数据特性

目前有许多研究主要从数据量的角度考察和评价大数据，〔26〕但实际上，大数据并非简单地指向规模化的数据量，而是指向认识论与方法论上的重要变革。在认知范式转变的意义上，大数据具有以下四个核心特性：

首先，大数据是行为而非行为客体。〔27〕仅存在超大体量的原始数据而不附加后续的管理与分析处理，不能称之为大数据，最多只能称之为“大数据库”。大数据呈现的事实或规律并非自始以完整形态存在，而是随着数据挖掘、分析的不断深入而逐渐成型。目前在刑事司法裁判领域出现的以“镶嵌论”（mosaic theory）为代表的事实认知模式，正是对大数据这一特性的直接反映。该理论认为，分散的数据碎片尽管对其占有人价值有限，但将这些碎片通过特定模型组合起来，则会产生不可估量的整体价值。〔28〕

其次，大数据引发认知范式转变基于一个基本前提，即假设特定主体的行为或偏好存在相对稳定的模式。从该假设出发，大数据所做的是通过积累和分析海量的“数据足迹”以发现目标对象的运行趋势，并以此为基础激活相应的解释、监控、预测、规划等机制。〔29〕这种假定存在某种行为模式并试图通过大量数据计算以发现该模式的思维，意味着传统的预先设定问题、再进行数据分析的思路难以为继，取而代之的是将数据分析前置并在分析中逐渐发现问题的思维过程。〔30〕

再次，大数据采用的是“数据—理论模型—特定现象”的认知范式，即以数据挖掘为分析起点，在此基础上概括出一般模型，最终以包括可视化等在内的方式实现精准定位或作业，这恰与传统的“特定现象—理论模型—数据”的认知范式相反。

最后，大数据的后续应用目标与前期数据收集之间难以确保精准对应。就大数据而言，对半结构化或非结构化数据的前期搜集和存储尤为关键，其中特别需要注意的是对动态数据和碎片数据的及时固定或实时处理。从另一个角度看，这一特征也意味着在数据处理过程中“数据噪音”增多，而数据分析最终要适用于特定主体或情形，因此，如何在即时性与精准性之间求得平衡，是大数据给数据处理技术和应用带来的新挑战。

基于大数据的基本特性，以及由此带来的认知范式转变，犯罪治理活动也呈现出一些新的重要变化：一是服务于刑事侦查活动的数据存留及共享义务开始大幅度扩张；二是以

〔26〕 例如在欧盟委员会（European Commission）2013 年发布的有关大数据的纪要中，就以世界每年产生的数据量作为“大数据”概念的核心。See European Commission, *Factsheet: What Is Big Data*, issued December 7, 2013, available at http://europa.eu/rapid/press-release_MEMO-13-965_en.htm, visited on December 20, 2016.

〔27〕 也有学者将大数据定义为过程，这与本文此处的观点可谓异曲同工。See James R. Kalyvas and David R. Albertson, *A Big Data Primer for Executives*, in James R. Kalyvas and Michael R. Overly (eds.), *Big Data: A Business and Legal Guide*, Boca Raton: CRC Press, 2015, p. 1.

〔28〕 See David E. Pozen, *The Mosaic Theory, National Security, and the Freedom of Information Act*, 115 The Yale Law Journal 628 (2005).

〔29〕 参见孟小峰、慈祥：《大数据管理：概念、技术与挑战》，《计算机研究与发展》2013 年第 1 期，第 148 页。

〔30〕 See Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 Washington Law Review 40 (2014).

犯罪风险防控为指导的犯罪治理活动启动时点的前移。

（二）服务于刑事侦查的数据存留及共享义务的扩张

大数据的具体运用依赖规模化的数据挖掘。数据碎片经由大规模搜集、重组、认知之后形成有效信息，而这一模式发挥效用的前提是数据搜集的常规化、普遍化。基于此，大数据介入犯罪治理通常引发三种相互联系的制度设计：其一是明确特定主体的数据存留、共享、披露义务；其二是鉴于大规模数据搜集可能引发与公民基本权利的冲突，设立以个人信息权为核心的基本权利保障机制；其三是为实现犯罪治理目的，形成刑事司法领域个人信息保护的例外规定。以网络安全法为例，首先，网络安全法第24条明确了特定业务中网络运营者收集用户真实身份信息的义务；其次，基于个人信息保护诉求，第40条至第45条就个人信息的收集、使用等行为，设定了条件和边界；再次，为实现犯罪治理目的，第28条进一步规定，“网络运营者应当为公安机关、国家安全机关依法维护国家安全和侦查犯罪的活动提供技术支持和协助”。从以上相关法律规定可以看出，在大数据导致数据存留及披露义务一般化的大背景下，立法基于个人信息保护目的而设定的限制性规定，主要指向包括网络服务提供商在内的私主体。而对于公安司法机关，只要符合法律规定的合理目的，在信息准入层面公安司法机关基本不存在获取信息的实质性障碍，其职责主要是在获取信息后履行保密、保管、专用以及销毁等义务。

事实上，权力机关之间以提升社会管理职能为目的的信息收集与共享平台，正在逐步建立。例如，根据2016年国务院《政务信息资源共享管理暂行办法》第2条，各部门之间信息交流与共享的适用范围，被扩展至“政务部门直接或通过第三方依法采集的、依法授权管理的和因履行职责需要依托政务信息系统形成的信息资源等”。结合2010年中纪委、中组部、中宣部等联合发布的《关于建立和完善执行联动机制若干问题的意见》、2011年公安部等十五部门联合发布的《关于建立实名制信息快速查询协作执法机制的实施意见》、2014年最高人民法院、证监会联合发布的《关于加强信用信息共享及司法协助机制建设的通知》、2016年最高人民法院、公安部联合发布的《关于建立快速查询信息共享及网络执行查控协作工作机制的意见》等规范性文件，可以更加清晰地看出国家权力机关之间信息共享的大趋势。在此规范框架下，数据仓库建设、联机分析以及数据挖掘形成了一整套犯罪治理模型，并且，随着大数据的进一步推广，其很可能在未来成为犯罪防控与侦查的核心手段。〔31〕

（三）基于风险防控的犯罪治理活动启动时点的前移

基于风险防控理念，犯罪治理活动的启动时点不断前移，这一方面表现为以预测警务（predictive policing）为代表的一般犯罪防控，另一方面体现为具体案件中侦查活动的提前启动。

预测警务一般被定义为基于相关数据采用量化分析等技术，协助警察识别犯罪风险并进行犯罪预防或犯罪治理等活动。〔32〕其有以下三项主要特征：（1）以识别和预测犯罪风险为主要目的；（2）以大数据分析为主要模型；（3）预测结果将触发相应的犯罪控制措施。依据功能差异，可将预测警务区分为四种类型：预测犯罪活动、预测潜在犯罪人、预测犯

〔31〕参见张兆瑞《关于公安大数据建设的战略思考》，《中国人民公安大学学报》2014年第4期，第19页。

〔32〕See Walter L. Perry *et al.*, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, RAND Corporation, 2013, p. xiii; Rutger Rienks, *Predictive Policing: Taking a Chance for a Safer Future*, Korpsmedia, 2015, p. 19.

罪人身份、预测犯罪被害人。〔33〕其中，预测潜在犯罪人和预测犯罪人身份均指向可能实施犯罪行为或具有犯罪嫌疑的特定个人；其间的区别在于，前者用于预测已知个人未来可能实施的特定犯罪行为，后者则用于预测已知犯罪可能对应的有特定犯罪记录的个人。无论基于何种模式，预测警务均要求执法人员以数据分析结果为依据，将人与犯罪行为进行预判性匹配，并基于该预判采取针对特定个人的犯罪控制措施。在此基础上，“犯罪地图”技术也同步发展起来，即通过强化监控与数据分析形成特定地区的犯罪热点可视化地图，从而引导警力分配和公民日常活动。〔34〕通过将犯罪治理活动在时间轴上前移，预测警务被认为能有效提高犯罪防控效率。

我国的犯罪治理同样关注到了预测警务的应用。2015 年，中共中央办公厅和国务院办公厅联合印发了《关于加强社会治安防控体系建设的意见》，强调通过“强化信息资源深度整合应用，充分运用现代信息技术，增强主动预防和打击犯罪的能力”。基于该工作思路，各地纷纷开始建立或强化预测警务系统，例如北京市怀柔区的犯罪数据分析和趋势预测系统、江苏省苏州市的犯罪警情预测系统、四川省推动的“雪亮工程”公共安全视频监控建设联网应用、江西省特殊人群大数据平台等。〔35〕预测警务系统使当地公安机关可以在犯罪高危地区提前布控，从而实现针对犯罪活动的有效精准打击。目前国内通过大数据进行侦查主要有两种模式：第一种模式是预测犯罪高发区。如江苏省苏州市自 2013 年起推行犯罪警情预测系统，该系统收录了十年来苏州市 1300 余万条警情数据和 7.8 亿条商铺信息，为科学配置警力、提速应急反应奠定了技术基础。〔36〕第二种模式是数据比对预测犯罪嫌疑人。例如，自本世纪初我国开始研发并逐步推广全国公安机关 DNA 数据库应用系统，至 2015 年 9 月，该数据库已经收录近 4000 万条 DNA 信息，〔37〕为个人信息比对和案件侦破提供了强大助力。

除预测警务外，犯罪治理活动启动时点的前移还体现在具体案件的侦查上。首先，通过应用第三方数据记录、大型数据库、预测性分析等手段，刑事诉讼启动标准的实现方式可能发生重大转变。〔38〕我国目前的立案标准仍是“有犯罪事实需要追究刑事责任”，而应用大数据技术的关键问题是，基于个人信息数据库分析而作出的行为预测，能否被用于犯罪事实补强或犯罪嫌疑人的确定。其次，侦查活动启动时点的前移，还体现在对立案前证据收集行为的效力确认。2016 年“两高一部”出台《关于办理刑事案件收集提取和审查判断电子数据若干问题的规定》（下文简称“刑事案件电子数据规定”），其中明确规定初查过程中收集、提取的电子数据可以作为证据使用。在大数据的背景下，数据收集与分析先于侦查人员对犯罪事实及刑事责任的认知而进行，这或许将成为常态；强制侦查措施先于

〔33〕 参见前引〔32〕，Perry 等书，第 xiv 页以下。

〔34〕 See Peter K. Manning, *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control*, New York University Press, 2008, pp. 17 – 19.

〔35〕 相关报道参见《互联网技术带给中国社会治安“全新可能性”》，http://www.mps.gov.cn/n2255079/n5137689/n5512386/n5512398/c5520647/content.html，2017 年 1 月 3 日访问。

〔36〕 同上。

〔37〕 参见刘烁：《全面深化公安机关 DNA 数据库建设发展应用，切实提升精确打击犯罪能力和服务实战水平》，《刑事技术》2016 年第 1 期，第 1 页。

〔38〕 See Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 (2) University of Pennsylvania Law Review 327 – 410 (2015).

立案行使的情况恐怕难以避免;〔39〕而在大数据形成和应用的过程中,对公民个人信息的干预恐怕也难以避免。

三、个人信息大数据与刑事正当程序的冲突

个人信息大数据在服务于查明事实这一实体性刑事司法价值的同时,其与以正当程序为依托的基本权利保障之间的紧张关系,尚未引起足够关注。基于当前大数据在犯罪治理领域的主要运用模式,这种紧张关系主要体现在三个方面:与无罪推定原则的冲突、与控辩平等原则的冲突、与权力专属原则的冲突。

(一)犯罪治理活动提前启动与无罪推定原则的冲突

如前所述,个人信息大数据介入刑事司法导致犯罪治理活动提前启动,一方面,犯罪预测将应对犯罪的模式由被动转为主动,由事后打击转向事前预防;〔40〕另一方面,在案件调查取证的过程中,侦查权得到扩张。基于这两方面的原因,产生了犯罪治理过程中个人信息大数据与无罪推定原则之间的冲突。2012 年刑事诉讼法第 12 条规定,"未经人民法院依法判决,对任何人都不得确定有罪"。这是无罪推定原则在我国法律中的体现。无罪推定原则的要义在于,对审前预判以及基于该预判进行的干预或限制公民基本权利的行为进行严格限制,这种限制与审前具有社会防卫性质的措施存在紧张关系;其核心问题是,基于风险考量,对尚未形成犯罪嫌疑的特定主体,是否以及在何种程度上可以对其基本权利进行干预或限制。〔41〕自"9·11"事件之后,这种紧张关系在打击恐怖主义犯罪等严重威胁国家及公共安全的犯罪方面,呈现出向强化犯罪控制方向倾斜的趋势。〔42〕在我国,这种趋势也反映在反恐怖主义法第五章关于"调查"的规定和国家安全法第 28 条关于反恐工作的规定之中:基于可能存在的恐怖主义犯罪风险,公安机关可以采取盘问、检查、传唤、提取或采集人体生物识别信息和生物样本、询问、收集调取相关信息和材料、查询、扣押、冻结财产、不超过三个月的约束措施等调查措施。

个人信息大数据在犯罪治理中的应用使得上述冲突开始向一般犯罪案件扩张。仍以预测警务为例,北京市怀柔区警方的预测系统目前主要针对盗窃类案件;江苏省苏州市的犯罪预测系统 2015 年就已经覆盖 91 种违法犯罪行为,〔43〕2016 年又搭建起非法集资预测预警处置平台;〔44〕江西省的特殊人群大数据平台针对的是包括服刑人员、刑满释放人员、戒毒

〔39〕参见龙宗智:《寻求有效取证与保证权利的平衡——评"两高一部"电子数据证据规定》,《法学》2016 年第 11 期,第 8 页。

〔40〕See Ian Kerr and Jessica Earle, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, in Mireille Hildebrandt and Katja de Vires (eds.), *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Routledge, 2013, p. 91.

〔41〕参见罗海敏:《预防性羁押的争议与适用》,《国家检察官学院学报》2012 年第 4 期,第 97 页以下。

〔42〕See Elies Van Sliedregt, *A Contemporary Reflection on the Presumption of Innocence*, 80 Revue Internationale de Droit Penal 247 - 248 (2009); 顾元:《美国总统战争权力的扩张与当代反恐战争中的人权保护》,《中国政法大学学报》2013 年第 8 期。

〔43〕参见《苏州公安 2015 年处理警情 385 万起,抓获逃犯 2948 名》,http://js.qq.com/a/20160113/028385.htm,2017 年 1 月 6 日访问。

〔44〕参见《平安苏州:新起点上再升级》,http://www.suzhou.gov.cn/news/szxw/201701/t20170104_833639.shtml,2017 年 1 月 6 日访问。

人员和社区矫正人员在内的 47 万余人。〔45〕

犯罪预测以及由此触发的后续治理措施在目标人员和目标犯罪这两个层面的一般化趋势，意味着犯罪风险防控的预防性措施与无罪推定原则的冲突已经由例外成为常规。此外，大数据在助力犯罪侦查并补强立案依据的同时，还会在一定程度上强化有罪推定的假设。〔46〕侦查人员在案发之初凭借经验和逻辑推理形成假设，以用于指引取证活动，并根据搜集到的证据碎片进行反复修正，最终形成完整的排除内在矛盾的证据链条，这种做法符合一般的侦查逻辑。但目前大数据多用于建立或强化有罪链条，执法人员在运用大数据之前已经形成的思维倾向，在经过基于碎片化信息重组的大数据分析之后，会被进一步强化和合理化。〔47〕

（二）大数据调查取证与控辩平等原则的冲突

从上文的分析可以看出，个人信息大数据发挥功效需要具备两个前提：一是具备规模数据的积累或获取能力；二是具备与该数据量相适应的分析能力。目前，数据收集以及原始数据库的形成，主要由政府部门、网络平台以及大型研究机构等承担；数据共享则主要发生在政府部门之间以及政府部门与商业机构之间，由此引发个人信息大数据与控辩平等原则的冲突。

1. 基于数据获取能力差异而形成的控辩不平等

刑事诉讼之举证责任在控方，为保障有效行使辩护权，辩方也被赋予一定的取证途径。2011 年《律师办理电子数据证据业务操作指引》规定了律师搜集和提取电子数据证据的四种主要方式：指导当事人取证、自行取证、申请包括司法行政机关在内的有权机关取证、请求包括网络服务提供商在内的第三方取证。在大数据介入刑事司法的背景下，后两种途径的重要性日益凸显，但目前尚未有规范性文件明确规定个人或其委托的律师如何以这两种方式从大数据的占有机构获取与案件相关的数据。同时，基于信息泄露、系统干扰、数据篡改等信息安全方面的考量，相关制度、规则在不同程度上都会对数据准入设置障碍，从而可能进一步扩大控辩双方在数据获取方面的能力差距。具体而言，这些障碍主要包括：基于国家安全考量而设置的限制，基于个人信息保护考量而设置的限制。

第一，就国家保密特权与数据库开放之间的关系而言，辩方在获取有利于被指控人的信息方面存在困难。以保守国家秘密法、刑法、网络安全法、反恐怖主义法等法律为框架，我国已经形成一整套国家保密机制。而相关法律和规范性文件一般着重强调有关单位的保密义务，对相关权利人的知悉权则缺乏明确规定。例如，政府信息公开条例第 14 条规定，行政机关不得公开涉及国家秘密、商业秘密、个人隐私的政府信息，其例外则包括权利人同意公开以及不公开可能对公共利益造成重大影响等两种情形，并且例外情形仅针对商业秘密与个人隐私信息。与此同时，保守国家秘密法第 9 条明确将“维护国家安全活动和追查刑事犯罪中的秘密事项”列入国家秘密的范畴。通过将刑事追查活动的特定信息列为国家秘密，使权力机关原则上不得开放相关大数据的进入、提取、分析、使用，辩方基于辩

〔45〕 参见《江西建成中国首个特殊人群大数据平台，覆盖 47 万余人》，http://money.163.com/16/1013/15/C393QD2L002580S6.html，2017 年 1 月 6 日访问。

〔46〕 See Joshua Fairfield and Erik Luna, *Digital Innocence*, 99 Cornell Law Review 981 – 1076 (2014).

〔47〕 See Sharad Goel *et al.*, *Combatting Police Discrimination in the Age of Big Data*, New Criminal Law Review, 2016, available at http://www.rshroff.com/uploads/6/2/3/5/62359383/policing – the – police.pdf, visited on Feb. 9, 2017.

护权能否以及如何从有关部门获取相关数据，法律和规范性文件的规定尚不明确。同时，在大力推进政府部门间信息共享的背景下，侦查机关的取证能力借助与其他政府部门的信息共享以及以网络运营商和网络服务提供商为主的商事主体的信息披露义务而大幅提升，这与辩方面临的取证限制形成鲜明对比。

第二，个人信息保护同样能够成为辩方取证的障碍。以网络安全法为例，其第40条规定了网络运营者的用户信息严格保密义务；第42条则以信息被收集者的同意作为网络运营者向他人提供个人信息的前提条件。如前所述，这种保密义务并非绝对，其例外情形主要限于第28条规定的公安机关、国家安全机关依法维护国家安全和侦查犯罪的活动。与这一例外相对应，2012年刑事诉讼法第52条规定，公检法机关在收集、调取证据时，对涉及国家秘密、商业秘密、个人隐私的证据，应当保密。但是，在大数据的背景下，个人信息碎片化愈发严重，以是否侵犯个人权利为标准划定取证行为边界的传统规则，已经难以在事前有效制约公权力的行使。〔48〕换言之，通过运用大数据技术，任何信息碎片都有可能经过重组而转化为涉及个人隐私的信息。大数据的这一特性，一方面导致信息被收集者的同意例外变得模糊不清和难以适用；另一方面亦使任何信息都有可能基于2012年刑事诉讼法第52条的规定被列入保密范畴，从而构成包括辩方在内的主体获取相关证据的制度障碍。

2. 基于数据分析能力差异而形成的控辩不平等

控辩双方在数据分析能力方面的差异尚未引起我国立法与司法实践的充分关注，但近期一些热点案件已经或多或少触及这一问题。以快播案为例，针对作为本案关键证据的四台服务器，司法机关实施了一系列数据提取和分析活动，例如委托鉴定中心筛选服务器远程访问IP地址，检验视频格式文件修改痕迹，提取29841个视频文件并认定21251个淫秽视频等。一方面，以上数据提取和分析活动需要大量专业技术人员的参与以及相应的物力、财力支持，以单个或少数几个律师进行辩护的传统策略已难以从证据内容本身进行有效应对；另一方面，相对于传统证据类型，电子数据证据的真实性或完整性在很大程度上依赖于取证过程的规范性和科学性，而在当前的机制下，辩方很难有效参与并监督取证过程。

已有域外学者指出，利用数据的庞大体量进行审前证据交换，已经演变成一种通过抬高诉讼成本来增强己方谈判筹码的诉讼策略。〔49〕在美国2009年的斯基林案中，〔50〕法院认为，通过向被告方开放特定数据库，政府已经充分履行证据开示义务，而不论数据库包含的数据量是否已经庞大到被告方不可能有效查找和提取文件。有学者将这种只提供数据库入口或最终证据，却不考虑控辩双方实际的数据分析能力和实质的程序参与的做法称为“文件倾倒”，并认为其构成控辩双方的实质不平等。〔51〕即便控辩双方掌握了同样的原始数据，但由于大数据的碎片化特征，双方也有可能以同样的素材拼组出截然不同的“事实”。可以预见的是，随着以信息碎片为核心要素的大数据不断深度介入刑事司法，数据本身的

〔48〕 参见裴炜：《比例原则视域下电子侦查取证程序性规则构建》，《环球法律评论》2017年第1期，第80页以下。

〔49〕 See Leah M. Wolfe, “*The Perfect is the Enemy of the Good*”: *the Case for Proportionality Rules instead of Guidelines in Civil E – discovery*, 43 Capital University Law Review 159 – 161 (2015).

〔50〕 United States v. Skilling, 554 F. 3d 529, 577 (5th Cir. 2009).

〔51〕 See Brandon L. Garrett, *Big Data and Due Process*, University of Virginia School of Law, Public Law and Legal Theory Research Paper Series 2014 – 45, available at http://ssrn.com/abstract=2481078, visited on January 10, 2017.

质量以及算法的可靠性，将成为大数据应用的关键。如何赋予辩方挑战控方分析方法与结论的能力，以及如何在控辩双方的解读之间以相对中立的方式进行评价，是大数据带给犯罪治理的新挑战。

有鉴于此，无论是在数据搜集、获取层面，还是在数据分析层面，大数据的应用都会使控辩双方的能力差异日益扩大。沿着社会数据化搭建起来的逻辑链条进一步分析，我们可以作出以下推断：基于形式平等的审前取证与庭审质证而建立起来的刑事诉讼规则将难以为继；如果刑事司法未来计划强化而非放弃控辩平等原则，就必须改革现有规则，以实现控辩双方在面对包括大数据在内的电子数据证据时，在数据搜集、获取、分析、呈现等方面具有大致相当的能力。

（三）司法外主体介入与权力专属原则的冲突

在大数据的背景下，谁掌握数据源，谁就有可能成为权力的实际执行者。人们在日常生活中遗留的大量数据痕迹并不会自动全面、及时地被录入政府数据库，而是会遗留在各类存储介质中，由包括网络平台在内的主体搜集并形成具有商业价值的数据分析基础。而大数据引发的数据存储及披露义务的扩张，以及犯罪治理活动启动时点的前移，意味着需要司法机关以外的社会主体介入犯罪治理，从而产生大数据侦查与权力专属原则的冲突。

2012 年刑事诉讼法第 3 条规定，除法律特别规定的以外，由公安机关、检察机关和法院行使特定的刑事司法权力，其他任何机关、团体和个人都无权行使。立法者之所以将这些权力归属于特定主体，一是以职权特定化防止权力扩张，二是以分工明确化来落实分权制衡，三是以程序法定化来避免法外行权。但是，在大数据的背景下，司法权特别是侦查权被稀释，侦控审之间的界限由于公众的普遍参与而有模糊化的趋势，由民众自主发起的包括监控、追查、公布、谴责等在内的犯罪治理活动在现有法律框架下也难以得到有效规制。

上述冲突的直接体现是“众包侦查”的出现。该模式以大规模数据积累为基础，但在运行逻辑上强调的是基于不确定数据来源的数据搜集分析。〔52〕如下图所示：

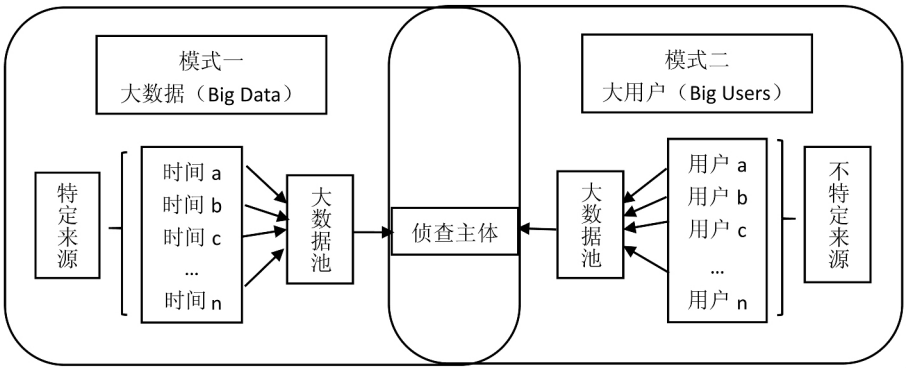

图 1 大用户模式（众包）与大数据模式的对比

〔52〕 关于大数据众包模式运行机理的分析，参见 Amin Ranj Rar & Muthucumaru Maheswaran，*Confidentiality and Integrity in Crowdsourcing Systems*，Springer，2014，pp. 1 – 4。

在众包模式下，监督和评估潜在犯罪行为，为实际发生的案件提供线索和证据的工作，开始由传统意义的执法机关向不特定社会群体转移。这种任务转移不仅针对已经发生的特定案件，还被广泛用于犯罪预测等活动。即便针对已经发生的案件，众包模式下相关参与群体的活动也区别于传统意义上被动提供证据或线索的证人。这些群体会主动向特定算法提供数据，后者会自动生成分析结果。借助众包模式进行犯罪风险治理的典型实践是2006年美国得克萨斯州采用的虚拟边境观测系统，该系统允许个人通过网络摄像头观测并实时汇报违规穿越美国与墨西哥边境的行为。〔53〕就我国而言，众包式执法表现为有明显公私合作性质的模式。同时，大型网络平台也逐渐成为司法运行的常规合作方。例如，2015年，浙江省高级人民法院与阿里巴巴合作，以淘宝搜集的用户收货地址作为司法文书送达地址。〔54〕

个人信息大数据介入犯罪治理所引发的上述三个层面的冲突，势必要求立法者对现有规则进行调整。从“权力—权利”二元互动出发，这种调整一方面需要秉持刑事正当程序背后的基本价值，从而维持国家行使刑罚权的合理边界和正当性基础；另一方面则需要依托大数据的特性，直面这种特性对现有规则形成的冲击。

四、基于无罪推定原则的规则修正

个人信息大数据对于无罪推定原则的冲击主要体现在两个方面：一是犯罪治理活动启动时点前移的普遍化，二是执法人员有罪推定倾向的强化。两者在侧重上有所区别，可以采用的对策也应加以区分。

（一）犯罪治理活动启动时点的前移

大数据介入犯罪治理导致以干预基本权利为基本特征的刑罚权提前启动，应对这一现象的关键在于，通过细化的程序设计，将可能干预包括个人信息权在内的基本权利的措施控制在合理范围内，并理清前期干预行为与后期刑事裁判之间的关系。同时，刑罚权的提前启动意味着刑事司法与治安管理之间的界限将进一步模糊，在大数据的背景下理清两者的关系无疑具有现实必要性。关于干预基本权利的措施的适用范围，应强调比例原则的合理引入与应用，〔55〕具体包括以下四个方面：

首先，从比例原则的基本要求出发，对刑罚权的介入时点和阶段进行明确划分，以此在后续的“权力—权利”衡平过程中构建基本框架。结合2012年刑事诉讼法、相关司法解释以及“刑事案件电子数据规定”可以看出，侦查取证活动已经由立案阶段提前至行政执法阶段甚至更早，犯罪初查的证据效力也已经得到普遍确认。在此背景下，根据大数据在当前的实际应用，有必要将刑罚权的启动划分为三个阶段：预测警务阶段、犯罪初查阶段、犯罪侦查阶段。后两者主要以立案与否作为阶段划分标准，较为模糊的是预测警务阶段与初查阶段的区分时点。2015年最高人民法院、最高人民检察院、公安部《关于办理网络犯

〔53〕 关于该系统的介绍，参见 Gordon Hull，*Texas Virtual Border Watch*，available at https://clas-pages.uncc.edu/gordon-hull/case-studies/texas-virtual-border-watch/，visited on January 11，2017。

〔54〕 参见《法律文书无法送达？浙高院与阿里合作，直接寄到淘宝收货地址》，http://news.163.com/15/1124/17/B972D3TE00014AED.html，2017年1月11日访问。

〔55〕 参见前引〔48〕，裴炜文。

罪案件适用刑事诉讼程序若干问题的意见》（下文简称“办理网络犯罪案件意见”）规定的初查启动要件是“案件事实或者线索不明”以致无法判断是否达到犯罪追诉标准。据此，初查的启动至少应当具备具体的、特定的犯罪事实或线索，这意味着由概括性数据分析得出的类似“犯罪地图”的预测警务所获得的信息，不应成为启动初查的条件。

其次，需要依据预测警务、初查和侦查三个阶段划分各个阶段可以干预的个人信息类型，并在此基础上配置相应的调查取证措施。具体而言，隐私权仅是个人信息权的下位概念，对公民基本权利的保护不仅包括相对完整的电子记录，还应扩展至敏感性较低的“关于数据的数据”。对敏感类信息的干预仅限于立案侦查阶段；初查阶段以不干预个人敏感信息为原则，并以经特殊程序许可的有限干预为例外；预测警务阶段则严格禁止对个人敏感信息的干预。就可以配置的调查取证手段而言，“刑事案件电子数据规定”并未作限制性规定。《人民检察院刑事诉讼规则（试行）》（下文简称“刑事诉讼规则”）第 173 条规定，初查阶段只能采取“询问、查询、勘验、检查、鉴定、调取证据材料等不限制初查对象人身、财产权利的措施”，并明确将强制措施、查封、扣押、冻结初查对象的财产以及技术侦查排除在外。该规定尽管对比例原则有所体现，但仅针对检察机关。“办理网络犯罪案件意见”尽管对初查环节可以采用的侦查措施进行了限制，但该规定的缺陷是并未涉及技术侦查措施。〔56〕笔者认为，初查仅仅构成侦查活动的一个“引子”，其本身并无必要拓展为相对独立的环节。目前立法与司法解释对初查的规定也主要旨在定性，因此，应当维持对初查阶段调查取证活动的严格限制。如果需要采用的取证手段超出该限制，且该手段属于专业取证指南或规范规定的必需手段时，侦查机关若计划推进取证，则应当进入立案程序。

再次，就侦查而言，对于基于接受网络服务或日常工作、生活、学习等活动而由网络服务提供者、工作单位、就学机构、研究机构等搜集的个人信息大数据，法律应当设置高于政府数据库的数据接触门槛，并以正式立案为界限；对于预测警务和初查阶段，则确立以不得获取为原则、以有限披露为例外的基本规则。

最后，从证明力与证据能力的角度出发，一方面，可以考虑在立法中明确规定根据大数据分析得出的不利于被指控人的过往行为模式仅能做侦查线索使用，而不得作为证据使用，但是，有利于被指控人的大数据分析意见则可以经证据补强印证后作为定案证据使用。另一方面，通过制定司法解释和执法机关行为规范，明确各类数据所适用的事项、途径、手段、限度等。例如，针对执法合法性问题，应明确以执法记录仪的记录为准；没有按照规定配备执法记录仪或者相关记录灭失却无合理解释的，应作出不利于执法人员的推断。

（二）执法人员有罪推定倾向的强化

目前，大数据主要用于构建有罪或风险成立的论证链条，但其本身也可以成为证明当事人行为合法的依据。公安司法机关对后者已有所认知，但主要适用于执法人员而非被指控人。例如，2016 年《关于深化公安执法规范化建设的意见》就要求“建立健全执法全流程记录机制，全面推行现场执法活动视音频记录制度”；“刑事案件电子数据规定”第 14 条明确规定，在条件具备的情况下，应当对收集、提取电子数据的相关活动进行录像。这些规定本身有助于提高侦查人员调查取证活动的规范性，反之也可用于证明或否定其执法行

〔56〕 关于“两高一部”电子数据证据规定中初查问题的探讨，参见前引〔39〕，龙宗智文。

为的合法性。

基于上述考量，国际学界逐渐发展出了“数字无罪”（digital innocence）概念，以应对大数据背景下出现的数据挖掘偏见和数字证据选择性忽视等现实问题。〔57〕这一方面意味着需要在规则设计上向有利于被指控人的方向进行倾斜；另一方面则意味着基于数据存储与提取的跨部门、跨行业属性，这种规则变动需要通过综合立法的形式协调进行。

首先，就规则倾斜而言，新的规则需要对大数据拼组和重新演绎信息的属性形成明确认识，并强调这种“拼组”和“再演绎”本身就有可能构成对事实的合理怀疑，从而成为案件定罪量刑的障碍。之所以说会成为定罪量刑的障碍，是基于这样一项基本假设，即最终呈现于庭审并作为判断罪与非罪依据的证据链条，只是依赖逻辑与经验对以往事实的重组而非完整再现。证据规则上区分直接证据与间接证据、实物证据与言词证据，就反映出对这一特性的考量。实践中出现的电子数据证据关联性证明困难，也从侧面折射出碎片化信息在形成较为完整的证明链条的过程中所存在的问题。〔58〕

其次，就规则的跨部门、跨行业协调而言，立法者应当考虑基于当前三大诉讼法推行的电子数据证据立法规范，建立起刑事司法主体与其他政府部门在包括大数据在内的电子数据提取、分析、处理等活动中的衔接配合机制，其中需要特别关注基本权利受到干预的公民在以上活动中所享有的权利及其救济途径。

五、基于控辩平等原则的规则修正

就大数据应用与控辩平等原则的冲突而言，需要分别在数据获取能力与数据分析能力这两个层面，对控辩双方的力量对比加以调整。

（一）数据获取能力的平衡

就数据获取能力而言，立法的关键是明确在何种情况下、以何种方式、在何种程度上可以对辩方开放特定数据库或提供特定数据。相关制度建设应当区分两种情形：一是辩方知晓存在有利于本方的数据，二是控方知晓存在有利于辩方的数据。

首先，就第一种情形而言，关键在于如何在确保数据安全的前提下，尽可能保障辩方有效获取该数据并提交法庭审查。从数据安全的角度看，以立法的方式原则性地要求包括政府机构、市场主体等在内的大数据占有者向个人直接开放全部数据，并非理想的问题解决方式。从更为现实的角度看，较为合理的方式是，基于个案的特定线索或依据，基于相关权利人的申请，由特定机构代为收集和提取数据。

在控辩平等原则的框架下，这种制度设计至少应包含两项要求。其一是通过立法在原则上明确特定主体的配合义务，以及不履行该义务时的救济措施。其二是明确证据收集、提取程序的启动要件，以及整个过程中辩方的参与权、知情权及其保障措施。在设计制度时需要特别注意的是，由于大数据可能涉及国家机密、商业秘密与个人隐私，所以，在特

〔57〕 参见前引〔46〕，Fairfield 等文；James S. Liebman *et al.*，*The Evidence of Things not Seen: Non – matches as Evidence of Innocence*，Columbia Law School Public Law & Legal Theory Working Paper Group 2012，No. 13 – 333，available at https://papers.ssrn.com/sol3/papers.cfm? abstract_id = 2194117，visited on Jan. 17，2017。

〔58〕 参见刘品新 《电子证据的关联性》，《法学研究》2016 年第 6 期，第 176 页以下。

定机关的配合义务以及相应的收集、提取程序上，需要根据数据类型作区别处理。

基于此，立法者可以考虑引入专门的审查和令状制度。具体而言，可以考虑由相关权利人提出申请，由检察机关或法院对申请进行审查，并在条件符合时发布令状，交由侦查机关和占有或控制申请所涉大数据的第三方主体进行取证。针对任何特定类型的大数据，申请收集、提取的相关权利人应就以下两个基本事项提供初步的证明材料：一是该大数据为特定主体占有或控制，二是该大数据与案件具有关联性。在整个过程中检察机关应当发挥监督职能，在确保取证行为规范、合法的同时，保障特定数据在后续处理及使用过程中的安全。

其次，就第二种情形而言，关键在于如何确保控方在取证和举证过程中兼顾有利于辩方的证据。可以预见，控方掌握大数据证据材料将是更为常见的情形。而前文的分析已经表明，大数据证据材料容易强化侦查人员的有罪推定倾向。因此，就有利于辩方的信息而言，除了强化数据搜集和分析者的信息共享和开放义务以外，还需要适当强化控方对证明被指控人无罪或无违法犯罪风险的信息的搜集。这种强化可以从以下两个层面入手：一是在侦查取证环节引入第三方评价或裁判，就数据搜集与分析是否存在前置性偏见进行审查；二是明确启动干预活动对应信息的正面和负面清单，为司法、执法人员提供详细指引。这两个层面是相互联系的，当前者实施到一定阶段并积累了一系列案件经验之后，即可以规范性文件的形式形成针对特定大数据收集、提取的正负面清单。

（二）数据分析能力的平衡

强调数据获取能力的平衡主要针对的是有利于辩方的大数据证据，而数据分析能力的平衡则主要着眼于强化辩方针对不利于本方的证据的质证能力。从前文的分析可以看出，就可能作为刑罚权启动依据的大数据而言，其形成过程可能早于刑事诉讼程序的启动。因此，被指控人或权利受到干预的主体能否在这一阶段介入，或者获取该阶段证据材料所处理的相关信息，将直接影响本方的质证能力。从这个角度讲，在大数据领域，随着侦查活动启动时点的前移，基于控辩平等原则的要求，有必要将辩方的程序性介入也相应提前。由此进一步衍生出三项制度设计要求：第一，在条件允许的情况下，尽可能保证权利受到干预的主体或其代理人参与到取证的过程中来，或者至少应保证相关权利人的知情权。第二，应当明确参与诉前大数据收集、提取、分析等活动的人员的出庭作证义务。第三，控方在举证时需要对诉前搜集并用作证据使用的大数据证据及其搜集、分析、处理过程加以说明。

除在介入时间上的平等待遇，更重要的是在实体层面确保辩护权的有效实施。基于大数据自身的特性，同时从保障辩护权有效行使和庭审实质化的要求出发，首先有必要建立审前大数据证据开示制度，在保障辩方知情权的同时，确保辩方充分了解此类证据的收集、提取、分析、存储、使用等情况。从制度设计的角度看，可以考虑依托 2012 年刑事诉讼法第 182 条设立的庭前会议制度，由审判人员主持，由控辩双方就大数据证据材料进行开示，并在必要时引入有专门知识的人进行说明。

在此基础上，立法者有必要完善专家辅助人制度，以适应大数据介入犯罪治理的现实要求。2012 年刑事诉讼法第 192 条规定的专家辅助人制度，主要针对鉴定意见这一静态的证据类型。“刑事案件电子数据规定”第 21 条对此作了扩展，即规定控辩双方在展示电子数据时，亦可在必要时聘请具有专门知识的人进行操作，并就相关技术问题作出说明，从而将专家辅助人制度扩展至对电子证据的展示。这对于强调数据处理过程的大数据证据而

言，无疑是一个重要的进步。但是，结合之前的分析可以看出，过程性作为大数据证据的基本属性，其并非只有在需要当庭展示时才体现出来，而是贯穿数据收集至最终呈现的整个过程，而这种过程性本身会直接影响该类证据的真实性、关联性甚至合法性。从这个角度讲，就大数据证据而言，专家辅助人的作用不限于法庭展示，还应当涉及对审前程序中证据收集、分析过程的说明，无论该证据在提交法庭审查时是否需要借助多媒体设备出示、播放或者演示。此外，当大数据证据的提取早于刑事诉讼程序的启动时，彼时处理数据的相关人员不仅应当在特定情况下负有出庭作证或进行说明的义务，其证言也应当在必要时由专家辅助人进行说明。同时，从前面的分析可以看出，如果在庭前会议中针对包括大数据在内的电子数据证据进行交换，则可以考虑将专家辅助人制度也适用于该环节。这样一方面便于控辩双方更有效地理解大数据证据，另一方面也有助于防止采用“数据倾倒”的诉讼策略。

六、基于权力专属原则的规则修正

数据的资源化衍生出前文论及的社会治理权力的外溢。在犯罪治理领域，这种权力外溢不仅体现在司法机关与其他传统上不具有司法职能的机关、组织、个体之间，同时也体现在司法机关与以商业主体为代表的其他社会主体之间。侦查取证活动时点的前移、网络平台等主体的介入、众包式警务的出现等，这些现象无一不对传统的以公检法机关为核心的司法运行模式形成挑战。要应对这些挑战，尤其需要关注两个问题：一是非司法行政主体的参与方式，二是证据规则在立案前调查环节与侦查环节的衔接。

（一）非司法行政主体参与犯罪治理的方式

如前所述，无论是大型网络平台或其他社会机构所负担的数据存留及披露义务，还是众包式侦查，社会公众参与犯罪治理的方式已经因为大数据发生了深刻变化。相对于传统侦查活动中相关主体承担的配合义务，大数据介入刑事司法使这种配合义务由个案演变为一般性义务，从而在一定程度上使这些主体成为国家行使刑罚权的必然延伸。

从这个角度讲，基于以平衡国家权力与公民权利为核心的比例原则的要求，针对预测警务、初查与侦查这三个阶段，占有或控制大数据的社会主体在存留及披露相关数据的义务方面应当有所区别。其中，在侦查环节大数据占有主体的配合义务最高，初查环节次之，预测警务环节再次之。就数据存留义务而言，由于在收集和存留之初难以预判数据的后期用途，所以，其存留期限的划分不应以大数据后续可能适用的领域为标准，而应当根据大数据所涉个人信息的敏感程度来加以区分。〔59〕就数据披露义务而言，则应当将数据的敏感程度与预测警务、初查、侦查三个阶段进行匹配。其中，涉及个人隐私的大数据仅能在正式立案之后向侦查人员提供；在初查和预测警务阶段，则仅能针对一般个人信息进行调取和分析。

对于就数据占有而形成的社会主体所承担的犯罪治理义务，需要通过法律对其设置必要的限度。一方面，以司法行政机关为主导的犯罪数据库建设应当发挥主导功能，避免向包括网络平台在内的社会主体施加过重的配合义务。另一方面，在侦查层面，数据中介是否承担以及以何种方式承担配合义务，应基于个案进行判断，对该中介的行为授权、具体

〔59〕 参见裴炜 《犯罪侦查中网络服务提供商的信息披露义务》，《比较法研究》2016 年第 4 期，第 103 页。

的案因、侦查目标、对象、范围、期限、方式等要素进行审查。

（二）刑事立案前后的证据规则衔接

在司法实践中，立案前的证据调查活动已成通行做法；并且，大数据的应用使这一活动的启动时点不断前移，由此引发立案前后证据调查与侦查活动的衔接问题。如果在不受立案后的刑事诉讼程序对于取证行为的限制的情况下，过早地赋予在此期间获取的大数据以证据资格，则一方面可能因大数据的质量得不到保障而影响定罪量刑的合理性与准确性，另一方面则可能损及法院审查认定对审前环节的引导。

就初查阶段而言，目前相关法律和规范性文件对于在此期间取得的证据，原则上予以认可。但由于该证据的搜集发生在刑事诉讼程序正式启动之前，所以，犯罪嫌疑人的诉讼权利保障在这一阶段无法适用。根据《最高人民法院关于适用〈中华人民共和国刑事诉讼法〉的解释》第 65 条的规定，行政机关在行政执法和查办案件的过程中收集的电子数据作为定案根据需满足两个条件：一是法庭查证属实，二是收集程序合法。这两个条件尽管同样适用于立案后的取证活动，但对程序合法的解释则存在差异。以“刑事案件电子数据规定”第 7 条规定的“收集、提取电子数据，应当由二名以上侦查人员进行”为例，初查时相关人员在收集、提取大数据证据材料时如未遵守该规定，是否属于程序违法，从而影响该证据材料的证明力甚至是证据能力，以及建立在该证据基础上的案件事实认定是否因程序瑕疵而导致合理怀疑，对于这些问题缺乏明确规定。究其根本，这是因为行政程序与刑事程序的核心理念和基本运行逻辑不同，所以，在程序瑕疵对证据证明力的影响上，认定亦有差异。

从大数据的特性及现实应用的角度出发，笔者认为至少可以从以下几个方面入手构建初查与侦查阶段的证据衔接机制。首先，在条件允许的情况下，初查活动应尽可能保持与侦查活动在规则适用上的一致性，以确保此阶段取得的证据材料的质量向侦查阶段看齐。其次，在无法达到刑事诉讼规则对侦查行为的要求时，初查阶段取得的证据材料仍可作为证据使用，但非经实质性证据补强不得单独作为定案根据。最后，在上述第二种情况下，强化相关取证人员的出庭作证义务，并在原则上应当有专家辅助人对此类证据的真实性和关联性进行论证。

就预测警务而言，其基本原理是通过大数据分析发现特定的犯罪行为模式，由此引出证据衔接方面的一个关键问题：在案件进入刑事诉讼程序之后，该行为模式对于事实认定的效力如何。此类证据基于已经存在的记录或观察推断当前案件中犯罪嫌疑人与犯罪行为之间的联系，但这种推断存在两个主要问题：其一是逻辑层面的可靠性，其二是用于演算行为模式的计算模型的可靠性。〔60〕同时，相对于已经进入侦查程序的证据调查活动，预测警务活动由于缺少统一的政策与法律规范指引，在具体应用过程中缺少透明度和有效的外界监督，导致相关行为人的权责不明确，这会进一步降低行为模式预测的可靠性。〔61〕当

〔60〕 See David Robinson & Logan Koepke, *Stuck in a Pattern: Early Evidence on "Predictive Policing" and Civil Rights*, issued in August 2016, available at https://www.teamupturn.com/static/reports/2016/predictive-policing/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf, visited on Feb. 9, 2017.

〔61〕 参见彭知辉：《基于大数据的警务预测——局限性及其顺应之道》，《中国人民公安大学学报（社会科学版）》2016 年第 2 期，第 40 页以下。

然，这些预测会随着刑事诉讼程序的向前发展，而可能逐渐通过其他证据材料的补强形成切实可信的证据。但在此之前，对于依据大数据分析而推导出的行为模式，需要谨慎对待。原则上，基于大数据分析而推导出的行为模式，仅能作为犯罪侦查线索而非定案根据来使用。但是，在一些特殊案件中，例如性犯罪案件中，大数据分析所推导出的行为模式与传统意义上的品格证据具有一定的相似性，故可以考虑将其作为证据链条的一部分提交法庭审查。〔62〕同时，就这种情形，立法应当设置明确的证据补强规则，即非经其他实质性证据补强，大数据证据不得单独作为定案根据。

Abstract: The information revolution that the modern society is experiencing has led to profound transformation of the ideology and paradigm of state governance, which are particularly embodied in the interaction between state power and human rights. Against this background, the protection of personal information entails not only the protection of the information itself, but also the prevention of the abuse of the information, which may lead to the infringement upon other legal rights of citizens, either substantial or procedural. In the context of criminal justice, information revolution introduces personal information big data into the process of crime control. The key attributes of big data-process and algorithm reliance, behavioral pattern oriented prediction, data exploration based cognition paradigm, and data fragmentation-have led to the corresponding changes in the ideology and model of crime control. The changes can be observed mainly from two aspects: the expansion of third parties' obligations of personal data collection, retention, and data sharing on the one hand, and an earlier starting point of crime control for the purpose of risk management on the other. While personal information big data can contribute to the identification of criminal risks and, subsequently, to the optimization of the allocation of criminal justice resources, it can also get into violent conflicts with criminal justice due process, particularly with the principle of presumption of innocence, the principle of equality of arms between the prosecution and the defense, and the principle of legality regarding the allocation and monopoly of power. At this point in history, and from an information society perspective, an analytical framework elaborating the start point and the foothold of the reform of relevant procedural rules needs to be adopted to resolve these conflicts. The start point refers to the dynamic evolution of the "power-right" interaction in the digital era, while the foothold refers to the establishment of a new equilibrium between the dual value of criminal justice-that is, crime control on the one hand, and human rights protection on the other.

Key Words: personal information, big data, crime management, due process, principle of proportionality

〔62〕有学者建议，基于性犯罪的特殊性，我国应引入品格证据规则。参见王禄生《美国性品格证据适用规则之借鉴》，《法学》2014年第4期，第131页以下。